Blind Image Deblurring by Spectral Properties of Convolution Operators

Guangcan Liu[♯], Yi Ma[§,♯]

[§] Visual Computing Group, Microsoft Research Asia, Beijing, China

[♯] Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, USA

Abstract

In this paper, we study the problem of recovering a sharp version of a given blurry image when the blur kernel is unknown. Previous methods often introduce an image-independent regularizer (such as Gaussian or sparse priors) on the desired blur kernel. For the first time, this paper shows that the blurry image itself encodes rich information about the blur kernel. Such information can be found through analyzing and comparing how the spectrum of an image as a convolution operator changes before and after blurring. Our analysis leads to an effective convex regularizer on the blur kernel which depends only on the given blurry image. We show that the minimizer of this regularizer guarantees to give good approximation to the blur kernel if the original image is sharp enough. By combining this powerful regularizer with conventional image deblurring techniques, we show how we could significantly improve the deblurring results through simulations and experiments on real images, especially when the blur is large. In addition, our analysis and experiments help explaining why edges are good features for image deblurring.

I. Introduction

Figure 1 shows a very common result of a very blurry image of a car taken by a moving camera – or similar blurring effect can be observed in many surveillance photos where the camera is static but the car is moving. In many situations, we would like to recover the sharp version of the image so that details of the image (such as the numbers on the license plate) become recognizable again. It is in general impossible to correctly deblur the whole image if both the camera motion and the scene geometry are both entirely unknown. Nevertheless, if we only consider a small image region (e.g., the license plate area shown in Figure 1), it is reasonable to assume that the blurry image, denoted as $B \in \mathbb{R}^{n_1 \times n_2}$, is approximately generated by the convolution of a sharp image, denoted as I_0, and a *common* blur kernel, denoted as $K_0 \in \mathbb{R}^{m_1 \times m_2}$, [1]:

$$B \approx I_0 \otimes K_0, \quad \text{s.t.} \quad K_0 \in \mathcal{S},$$

where $\otimes$ denotes the discrete 2D convolution operator, S denotes the *simplex* of all possible blur kernels (i.e., nonnegative and sums to one), n_1, n_2 are the image sizes, and m_1, m_2 are the sizes of the blur kernel. So, the blind image deblurring problem is mathematically formulated as the problem of *blind deconvolution* [1], [2], [3], which is to recover the sharp image I_0 when the blur kernel K_0 is unknown [1].

The blind decovolution problem has been being investigated for several decades widely in optical society, image processing, computer graphics, and computer vision [1], [5], [6], [7], [8], [9]. Yet, as recently surveyed in [7], this problem is still far from being solved in the more general cases even with the notable progresses made recently (e.g., [10], [11], [12], [13]).

---

[1]*non-blind deconvolution* [4] is the restoration of I_0 when K_0 is given.
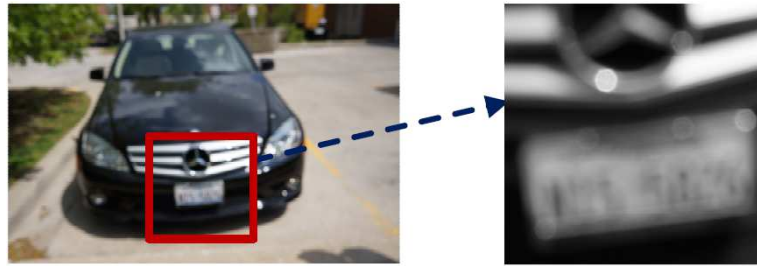

Fig. 1. **A scenario of image deblurring.** We take a picture of size 1624×2448 by using a NEX-5N camera. The picture is blurry due to an uncontrolled camera motion. We want to process the picture such that human eyes are able to recognize its details (e.g., the plate number).

A straightforward approach for blind deconvolution is to jointly seek the sharp image I_0 and the blur kernel K_0 by minimizing

$$\min_{I,K} \|B - I \otimes K\|_F^2, \quad \text{s.t.} \quad K \in \mathcal{S},$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. However, this problem is highly ill-conditioned, as it can be perfectly minimized by infinite number of pairs (I, K). For example, the *no-blur explanation* is a perfect solution: $K = \delta$ (a delta function) and $I = B$. Indeed, even if the blur kernel K_0 is given, the non-blind deconvolution problem can still be ill-conditioned [4], [14]. So, in general, it is necessary to regularize the desired solution for the image I:

$$\min_{I,K} \|B - I \otimes K\|_F^2 + \lambda f(I), \quad \text{s.t.} \quad K \in \mathcal{S}, \tag{1}$$

where $\lambda \geq 0$ is a parameter. The regularizer $f(I)$ is usually chosen as the *total variation* [15] or its variations [4], [14], [16]. However, as analyzed in [7], such image gradients based regularizer generally favors a blurry solution over a sharp one! Actually, whenever $f(I)$ is convex, it is easy to prove that (see the Appendix)

$$f(I \otimes K) \leq f(I), \quad \forall I \in \mathbb{R}^{n_1 \times n_2}, \ \forall K \in \mathcal{S}; \tag{2}$$

that is, minimizing $f(I)$ will encourage blurring, and thus the no-blur explanation (i.e., $I = B$ and $K = \delta$) is always favored by minimizing (1). Hence, it is critical to regularize the kernel K, resulting in the following optimization problem:

$$\min_{I,K} \|B - I \otimes K\|_F^2 + \lambda f(I) + \alpha h(K), \quad \text{s.t.} \quad K \in \mathcal{S}, \tag{3}$$

where $\lambda \geq 0, \alpha \geq 0$ are two parameters.

While it is obviously that the regularizer $h(K)$ plays important role in blind deconvolution, the existing proposals for $h(K)$ (e.g., the Gaussian prior $h(K) = \|K\|_F^2$ [6], [17] and the sparse prior $h(K) = \|K\|_1$ [13], [18]) have no significant effect on constraining the blur kernel [7]. Although such objectives can work well on some simple cases where there is no serious blur [2], they could not handle more difficult deblurring tasks such as images with a significant blur (e.g., Figure 1). Moreover, the Gaussian and sparse priors are inaccurate for characterizing real world blur kernel, which is usually the combination of a sparse curve-like kernel (which corresponds to the camera motion) and a dense Gaussian-like kernel (that models effects such as out of focus).

In this work, we derive a much more effective regularizer for the blur kernel that can significantly benefit the solution of the blind deconvolution problem. This regularizer is based on a simple but important observation about the spectral properties of an image as a convolution operator: For a given image (i.e., matrix), consider its convolution with any other image. The convolution defines a linear operator. *Then empirically, the spectrum (the set of eigenvalues) of this linear operator for a blurry image is significantly smaller than that for its sharp counter part (see Figure 2).* In fact, this can be proven to be true to some extend. Based on this observation, we devise a convex regularizer that tends to be minimized at the true blur kernel K_0. Namely, given an observed image B represented by a certain image feature $\mathcal{L}$, we deduce a convex function, denoted as $h^{\mathcal{L}(B)}(K)$:

$$h^{\mathcal{L}(B)}(K) : \mathbb{R}^{m_1 \times m_2} \to \mathbb{R}. \tag{4}$$

Unlike the previous work (e.g., [6], [13], [17], [18]) where the regularizer $h(K)$ is independent of the observed image B, our regularizer $h^{\mathcal{L}(B)}(K)$ explicitly depends on the given blurry image and encodes information about how the blurry image B is related to the sharp image I_0. It is hence somewhat natural to anticipate that this regularizer $h^{\mathcal{L}(B)}(K)$ would depend on the sharp image too. But rather surprisingly, as we will show, under fairly broad conditions, we can come up with a very effective regularizer $h^{\mathcal{L}(B)}(K)$ that does not depend on any information about the sharp image I_0 at all, and the desired kernel K_0 can be approximately retrieved by minimizing $h^{\mathcal{L}(B)}(K)$ directly.

Equipped with such new regularizer $h^{\mathcal{L}(B)}(K)$, for the blind deconvolution problem, we jointly seek the sharp image I_0 and the blur kernel K_0 by solving the optimization problem (3). Experimental performance of our algorithm is a bit surprising: Even when the observed image is blurred to the extent that human eyes cannot recognize its details (e.g., Figure 1), it is still possible for our algorithm to restore a sharp version with recognizable details. In addition to empirical evaluations, theoretical results presented in this paper could also help understanding the blind deconvolution problem:

- It is known that there are infinite number of ways to decompose a given blurry image B into the convolution of I and K [7]. Nevertheless, while I (or its edge map) is assumed to be very sharp, we show that such decomposition tends to be unique (or approximately so).
- Although it is widely observed and believed that edges are more important than smooth regions for image deblurring [10], [13], [17], [19], [18], there is still lack of good justification. To some extent, our analysis could explain why sharp edges are good features for deblurring.

---

[2] Such simple cases could be solved even without the regularizer $h(K)$. More precisely, by carefully choosing the initial solution and controlling the number of iterations, it is possible to obtain satisfactory results based on optimizing (1) [19].
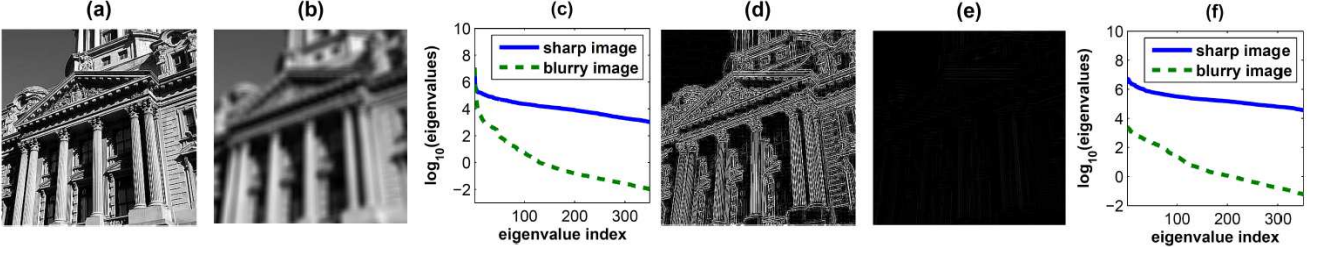
Fig. 2. **Example: blurring can significantly decreases the convolution eigenvalues of a sharp image**. (a) A sharp image. (b) The blurry image created by convoluting the sharp image with a Gaussian kernel. (c) Plots of all ($s_1 = s_2 = 18$) convolution eigenvalues of the sharp and blurry images, using $\mathcal{L} = \delta$. (d) The edges of the sharp image, using LoG. (d) The edges of the blurry image (note that the black area contains negative values). (e) Plots of convolution eigenvalues of the sharp and blurry images, choosing $\mathcal{L} = $ LoG.

II. BLIND DECONVOLUTION VIA SPECTRAL PROPERTIES OF IMAGE CONVOLUTION

In this section, we present the details for designing the regularizer $h^{\mathcal{L}(B)}(K)$. The final algorithm for blind deconvolution will be given at the end of this section.

A. Spectral Properties of an Image as a Convolution Operator

1) Preliminaries: The concept of *convolution* is well-known. We briefly introduce it for the ease of reading. Let X and Y are functions of two discrete variables (i.e., X and Y are matrices), then the formula for the 2D convolution of X and Y is

$$(X \otimes Y)(i,j) = \sum_{u,v} X(i - u, j - v) Y(u, v),$$

where $(\cdot)(i,j)$ denotes the (i,j)-th entry of a matrix. The convolution is linear and can be converted into matrix multiplication. Let $\nu(\cdot)$ be the vectorization of a matrix, then

$$\nu(X \otimes Y) = \mathcal{A}_{k_1,k_2}(X)\nu(Y), \quad \forall Y \in \mathbb{R}^{k_1 \times k_2}, \tag{5}$$

where $\mathcal{A}_{k_1,k_2}(\cdot)$ is the *Toeplitz matrix* [20] of a matrix, and k_1, k_2 are parameters. For a matrix $X \in \mathbb{R}^{l_1 \times l_2}$, its Toeplitz matrix $\mathcal{A}_{k_1,k_2}(X)$ is an $(l_1 + k_1 - 1)(l_2 + k_2 - 1)$-by-$k_1 k_2$ matrix.

2) Convolution Eigenvalues and Eigenvectors of Images: In this work, an image I is treated as a matrix associated with a certain feature filter $\mathcal{L}$:

$$\mathcal{L}(I) = \mathcal{L} \otimes I. \tag{6}$$

Typically choices for the feature filter $\mathcal{L}$ include $\mathcal{L} = \delta$ (i.e., using original pixel values as features) or $\mathcal{L} = $ "Laplacian of Gaussian (LoG)" (i.e., using edge features).

To formally characterize the procedure of transforming sharp images into blurry images, we define the so-called *convolution eigenvalues* and *convolution eigenvectors* as follows.

Definition 2.1: For an image $I \in \mathbb{R}^{l_1 \times l_2}$ represented by a feature filter $\mathcal{L}$, its first convolution eigenvalue, denoted as $\sigma_1^{\mathcal{L}}(I)$ or $\sigma_{max}^{\mathcal{L}}(I)$, is defined by

$$\sigma_1^{\mathcal{L}}(I) = \max_{X \in \mathbb{R}^{s_1 \times s_2}} \|\mathcal{L}(I) \otimes X\|_F, \quad \text{s.t.} \ \|X\|_F = 1,$$

where s_1, s_2 are called "sampling sizes" in this work. The maximizer to above problem is called the first convolution eigenvector, denoted as $\kappa_1^{\mathcal{L}}(I)$.

Similarly, the i-th ($i = 2, \ldots, s_1 s_2$) convolution eigenvalue $\sigma_i^{\mathcal{L}}(I)$ is defined by

$$\sigma_i^{\mathcal{L}}(I) = \max_{X \in \mathbb{R}^{s_1 \times s_2}} \|\mathcal{L}(I) \otimes X\|_F, \quad \text{s.t.} \ \|X\|_F = 1, \langle X, \kappa_j^{\mathcal{L}}(I) \rangle = 0, \forall j < i,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner production between two matrices. The maximizer to above problem is the i-th convolution eigenvector, denoted as $\kappa_i^{\mathcal{L}}(I)$.

In the following, we summarize some of their properties which will be used later:

- From (5), it can be seen that the convolution eigenvectors/values are exactly the right singular vectors/values of the Toeplitz matrix. So, for an image I with the associated Toeplitz matrix $\mathcal{A}_{s_1,s_2}(\mathcal{L}(I))$, its all $s_1 s_2$ convolution eigenvalues (and eigenvectors) can be found by computing the Singular Value Decomposition (SVD) of $(\mathcal{A}_{s_1,s_2}(\mathcal{L}(I)))^T \mathcal{A}_{s_1,s_2}(\mathcal{L}(I))$.

- Let $\sigma_{min}^{\mathcal{L}}(\cdot)$ denote the smallest (i.e., last) convolution eigenvalue of an image, then

$$\|\mathcal{L}(I) \otimes X\|_F \geq \sigma_{min}^{\mathcal{L}}(I) > 0, \forall \|X\|_F = 1. \tag{7}$$

- If $B = I \otimes K, K \in \mathcal{S}$, then (detailed proofs are in the appendix)

$$\sigma_i^{\mathcal{L}}(B) \leq \sigma_i^{\mathcal{L}}(I), \forall i = 1, \ldots, s_1 s_2; \tag{8}$$

that is, the blurring effects generally reduce the convolution eigenvalues of an image. When the original image is sharp, the reduction amount can be very significant, as exemplified in Figure 2. In particular, when the edge features are used, Figure 2(f) shows that it is even possible to have $\sigma_{min}^{\mathcal{L}}(I) > \sigma_{max}^{\mathcal{L}}(B)$.

The properties in inequality (8) provide a way to formally define the concept of *sharp image*, which appears frequently in the articles related to deconvolution: An image I is called τ-*sharp* if and only if $\sigma_{min}^{\mathcal{L}}(I) \geq \tau$, where $\tau > 0$ is a parameter.

B. A Convex Blur Kernel Regularizer

In this subsection, we derive a convex regularizer $h^{\mathcal{L}(B)}(K)$, which tends to have the minimal value at the desired blur kernel K_0.

1) Derivation: For ease of exploration, we begin with the "ideal" case that the blurry image B is exactly generated by the convolution of I_0 and K_0, i.e., there is no noise: $B = I_0 \otimes K_0$. As convolution operators are associative and commutative, the effect of feature extraction is

$$\mathcal{L}(B) = \mathcal{L}(I_0) \otimes K_0. \tag{9}$$

Here, $\mathcal{L}$ is a predefined feature filter. For the rest of this paper, we consistently choose $\mathcal{L} = \text{LoG}$, i.e., we use edge features by default (but our analysis applies to any convolution filters).

By Definition 2.1, we have that

$$\|\mathcal{L}(B) \otimes \kappa_i^{\mathcal{L}}(B)\|_F = \sigma_i^{\mathcal{L}}(B), \ \forall 1 \leq i \leq s_1 s_2,$$

where the sampling sizes s_1, s_2 are taken as parameters. By (9),

$$\|(\mathcal{L}(I_0) \otimes K_0) \otimes \kappa_i^{\mathcal{L}}(B)\|_F = \sigma_i^{\mathcal{L}}(B), \forall 1 \leq i \leq s_1 s_2.$$

Since the convolution operator is linear, we further have

$$\left\| \mathcal{L}(I_0) \otimes \frac{K_0 \otimes \kappa_i^{\mathcal{L}}(B)}{\|K_0 \otimes \kappa_i^{\mathcal{L}}(B)\|_F} \right\|_F = \frac{\sigma_i^{\mathcal{L}}(B)}{\|K_0 \otimes \kappa_i^{\mathcal{L}}(B)\|_F}.$$

Notice that $\left\| \frac{K_0 \otimes \kappa_i^{\mathcal{L}}(B)}{\|K_0 \otimes \kappa_i^{\mathcal{L}}(B)\|_F} \right\|_F = 1$. By (7), $\|\mathcal{L}(I_0) \otimes X\|_F \geq \sigma_{min}^{\mathcal{L}}(I_0), \forall \|X\|_F = 1$. Thus we have the following necessary conditions that constrain possible K_0:

$$\|K_0 \otimes \kappa_i^{\mathcal{L}}(B)\|_F \leq \frac{\sigma_i^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0)}, \ \forall 1 \leq i \leq s_1 s_2. \tag{10}$$

Define a regularizing function on the kernel K as

$$h^{\mathcal{L}(B)}(K) \doteq \sum_{i=1}^{s_1 s_2} \frac{\|K \otimes \kappa_i^{\mathcal{L}}(B)\|_F^2}{(\sigma_i^{\mathcal{L}}(B))^2}.$$

Then (10) implies that

$$h^{\mathcal{L}(B)}(K_0) \leq \frac{s_1 s_2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}. \tag{11}$$

Note that $\sigma_i^{\mathcal{L}}(B)$ could be significantly smaller than $\sigma_{min}^{\mathcal{L}}(I_0)$ (see Figure 2), and a randomly chosen blur kernel may not satisfy (10) or subsequently (11). Hence, the desired blur kernel K_0 should have relatively smaller value for $h^{\mathcal{L}(B)}(K)$. Or we could try to obtain an approximate estimate of the desired blur kernel K_0 by minimizing:

$$\hat{K}_0 = \arg\min_K h^{\mathcal{L}(B)}(K), \quad \text{s.t.} \quad K \in \mathcal{S}, \tag{12}$$

where S denotes the $(m_1 m_2 - 1)$-dimensional simplex, and m_1, m_2 are the sizes of the blur kernel.

It is easy to see that $h^{\mathcal{L}(B)}(K)$ is a quadratical (hence convex) function, namely $h^{\mathcal{L}(B)}(K) = (\nu(K))^T H \nu(K)$ with the Hessian matrix H given by

$$H = \sum_{i=1}^{s_1 s_2} \frac{(\mathcal{A}_{m_1, m_2}(\kappa_i^{\mathcal{L}}(B)))^T \mathcal{A}_{m_1, m_2}(\kappa_i^{\mathcal{L}}(B))}{(\sigma_i^{\mathcal{L}}(B))^2}, \tag{13}$$

where $\mathcal{A}_{m_1, m_2}(\kappa_i^{\mathcal{L}}(B))$ is the Toeplitz matrix of the i-th convolution eigenvector of B.

$\dfrac{\sigma^{L}_{max}(B)}{\sigma^{L}_{min}(I_0)}$: 0.7   0.9   1.5   1.2   1.4   1.0

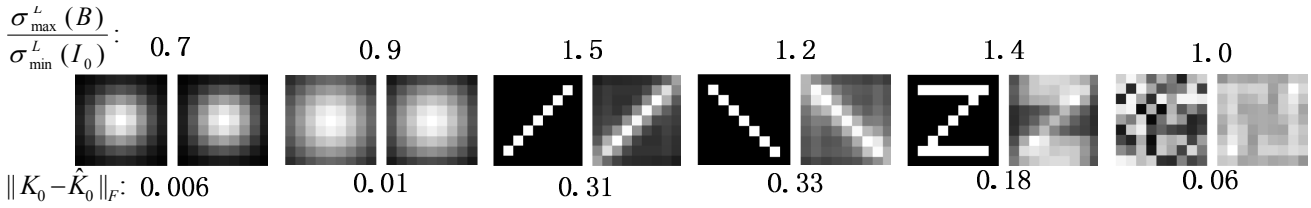$\|K_0 - \hat{K}_0\|_F$: 0.006   0.01   0.31   0.33   0.18   0.06

Fig. 3. **Demonstrating the effectiveness of the regularizer** $h^{\mathcal{L}(B)}(K)$, **using six synthetical examples.** Left: The blur kernel with size 9×9. Right: The kernel estimated by solving (12).

2) Analysis: We now study how effective the regularizer $h^{\mathcal{L}(B)}(K)$ is by evaluating how close its minimizer $\hat{K}_0$ is to the true kernel K_0 (proofs to all the theories can be found in the appendix).

Theorem 2.1: Suppose $B = I_0 \otimes K_0, K_0 \in \mathcal{S}$, and $I_0 \neq 0$. For the kernel $\hat{K}_0$ estimated by (12), we have that

$$\|K_0 - \hat{K}_0\|_F \leq \sqrt{2}\frac{\sigma^{\mathcal{L}}_{max}(B)}{\sigma^{\mathcal{L}}_{min}(I_0)}.$$

The above theorem illustrates that the edges are better than the raw pixels as a feature for the recovery of blur kernel, because the edge features can achieve a smaller bound for the estimate error (see Figure 2). This to some extent corroborates the previous observations (e.g., [10], [13], [17], [19], [18]) that the edge feature is a good choice for image deblurring. Note that in general $\sigma^{\mathcal{L}}_{min}(I_0)$ can be significantly larger than $\sigma^{\mathcal{L}}_{max}(B)$. For the example shown in Figure 2(f), we have computed the ratio $\sigma^{\mathcal{L}}_{max}(B)/\sigma^{\mathcal{L}}_{min}(I_0) = 0.15$, [3] then the above theorem suggests that the estimation error is upper bounded by 0.21. Figure 3 shows some more simulated examples of blur kernels directly estimated from a blurred Lena image, compared with the true kernels. Notice, that these kernels are obtained without any knowledge about the original image I_0 at all! This puts a strongly correct prior on the desired kernel entirely based on the given blurry image. These simulated results clearly demonstrate the effectiveness of the proposed regularizer $h^{\mathcal{L}(B)}(K)$.

Without any restrictions, it is known that there are infinite number of ways to decompose a blurry image B into the convolution of an image I and a blur kernel K [7]. Interestingly, when the image I (or its edge image) is assumed to be very sharp (i.e., the smallest convolution eigenvalue is large), Theorem 2.1 suggests that the decomposition for given image B tends to be unique in a way that the allowable kernels should be very close to each other. The following corollary makes this statement more precise: We call an image I is τ-sharp if $\sigma^{\mathcal{L}}_{min}(I) \geq \tau$ with $\tau = 2\sqrt{2}\sigma^{\mathcal{L}}_{max}(B)/\varepsilon$ for some small number ε. Then we have:

Corollary 2.1: Denote all possible τ-sharp decompositions of a blurry image B as $\Omega^{\varepsilon}_B = \{(K, I)|I \otimes K = B, K \in \mathcal{S}\}$. For any two pairs $(K_0', I_0'), (K_0'', I_0'') \in \Omega^{\varepsilon}_B$, we have that

$$\|K_0' - K_0''\|_F \leq \varepsilon,$$

where $\varepsilon > 0$ is any small parameter.

The above analysis are based on the assumption that there is no noise, i.e., $B = I_0 \otimes K_0$. Due to the fact the the blur in reality may not be uniform within an image, a more appropriate model is that $B = I_0 \otimes K_0 + N$, where N denotes unknown noise (errors). In this case, we have the following theorem to bound the estimate error.

Theorem 2.2: Suppose $B = I_0 \otimes K_0 + N, K_0 \in \mathcal{S}, \|\mathcal{L}(N)\|_F \leq \epsilon$, and $I_0 \neq 0$. For the kernel $\hat{K}_0$ estimated by (12), we have that

$$\|K_0 - \hat{K}_0\|_F \leq \sqrt{2}\frac{\sigma^{\mathcal{L}}_{max}(B) + ccond(B)\sqrt{s_1 s_2}\epsilon}{\sigma^{\mathcal{L}}_{min}(I_0)},$$

where $ccond(B) = \sigma^{\mathcal{L}}_{max}(B)/\sigma^{\mathcal{L}}_{min}(B)$ is the "convolution condition number" of B.

C. Practical Blind Deconvolution via Alternating Minimization

Theorem 2.2 illustrates that the estimation produced by (12) may not be so accurate with noise. Moreover, the estimated kernel by directly minimizing the regularizer is not always accurate enough even for noiseless cases (see Figure 3). To solve the overall blind deconvolution problem, we use the proposed regularizer $h^{\mathcal{L}(B)}(K)$ as an additional term to constrain the deconvolution problem. Namely, we jointly seek the sharp image I_0 and the blur kernel K_0 by minimizing the following objective function:

$$\min_{I,K} \|B - I \otimes K\|_F^2 + \lambda\|\nabla I\|_1 + \alpha h^{\mathcal{L}(B)}(K), \quad \text{s.t.} \quad K \in \mathcal{S},$$

[3]One may have noticed that the $\sigma^{\mathcal{L}}_{min}(I_0)$ shown in (10) (and from then on) actually refers to $\sigma^{\mathcal{L}}_{(s_1+m_1-1)(s_2+m_2-1)}(I_0)$ other than $\sigma^{\mathcal{L}}_{s_1 s_2}(I_0)$, because the convolution operator can change the sizes. The number 0.15 is calculated with considering this detail.

where we have chosen $f(I) = \|\nabla I\|_1$ (i.e., total variation) and $h(K) = h^{\mathcal{L}(B)}(K)$ in (3).

Although the above problem is nonconvex, the initialization step is no long crucial due to the strong image-dependent kernel regularizer: Unlike the previous algorithms (e.g., [6], [10], [13], [17], [18]) which need to carefully chose the initial solution, we simply choose the observed blurry image B to be the initial condition for I.

- While fixing the variable I, the blur kernel K is updated by solving $\min_K \|B - I \otimes K\|_F^2 + \alpha h^{\mathcal{L}(B)}(K)$, s.t. $K \in \mathcal{S}$, which is equal to the following quadratical programming:

$$\min_K \|\nu(B) - \mathcal{A}_{m_1,m_2}(I)\nu(K)\|^2 + \alpha(\nu(K))^T H \nu(K), \quad \text{s.t.} \quad K \in \mathcal{S},$$

  where the Hessian matrix H is computed by (13), and $\|\cdot\|$ is the ℓ_2-norm of a vector.
- While fixing the blur kernel K, the estimate of the image I is updated by

$$\min_I \|B - I \otimes K\|_F^2 + \lambda\|\nabla I\|_1,$$

  which can be solved by any of the many non-blind deconvolution algorithms developed in the literature. In this paper, we simply use the fast method introduced by [4].

Unlike previous blind deblurring methods that need to carefully control the number of iterations, we run the iterations until convergence. Usually, our algorithm needs about 100 iterations to converge.

On Choosing the Parameters. There are six parameters in total: The kernel sizes m_1, m_2, the sampling sizes s_1, s_2, and the trade-off parameters α, λ. Three of them, including m_1, m_2 and α, need be set carefully. The kernel sizes m_1, m_2 are better to be a bit larger than the true blur size, which requires trying the algorithm several times to determine. It also needs some efforts to choose the parameter α. When α is too small, our algorithm will always converge to the no-blur explanation (i.e., $I = B, K = \delta$); when α is set to be very large, the recovered image is very sharp, but unnatural. Usually, there exists such a threshold α^*: The optimal solution is always $(I = B, K = \delta)$ while $\alpha < \alpha^*$; the solution is often satisfactory while α is slightly larger than α^*.

After the kernel sizes being determined, we always set the sampling sizes as $s_1 = 1.5m_1$ and $s_2 = 1.5m_2$. The parameter λ has no significant influences. In our experiments, this parameter is chosen from the range of 0.001 to 0.002.

Computational Cost. Due to the developments of fast non-blind deconvolution (e.g., [4]), the procedure for updating the image variable I is already very fast. So we only analyze the costs for computing the Hessian matrix H and updating the kernel variable K. For simplicity, we assume that $s_1 = m_1, s_2 = m_2$. Then the complexity of computing the Hessian matrix H is $O(m_1^2 m_2^2 n_1 n_2 + m_1^3 m_2^3)$, where n_1, n_2 are image sizes. In each iteration, the Toeplitz matrix $\mathcal{A}_{m_1,m_2}(I)$ need be updated. So, the complexity of the blind deconvolution procedure is $O(n_s(m_1 m_2 n_1 n_2 + m_1^3 m_2^3))$, where n_s is the number of iterations needed to converge. Overall, our current algorithm is not fast while deal with large kernels. To process a 300×300 image with 13×13 kernel, our algorithm needs about one minute. While dealing with a 500×500 image with 31×31 kernel, the computational time increases to two hours. Nevertheless, it is possible to speed up the algorithm by fast Fourier transforms. We leave this as future work.

III. Experiments

A. Main Results

We test with five examples (two synthetical, three real): The synthetical images are the convolution of 300×300 natural images and 13×13 synthetical blur kernels; the real images are subregions (about 350×350) selected from large pictures captured by a NEX-5N camera (please refer to Figure 1). To show the advantages of the proposed method, we also test five state-of-the-art blind deconvolution algorithms, including [10], [17], [19], [18], and [13].

Figure 4 shows the comparison results. To save space, we only show the best result of the previous algorithms for comparision. On the simple examples with easy blur kernels and tiny noise level (the first three examples in Figure 4), our algorithm performs as well as the most effective baseline. While dealing with challenging cases (the last two examples in Figure 4), where the blur kernels are complicated, it can be seen that our algorithm works distinctly better than the most competitive baseline. This is because that our regularizer $h^{\mathcal{L}(B)}(K)$ contains strong and accurate priors about the blur kernels, whereas the kernel priors adopted by previous algorithms are too weak to handle such difficult deblurring tasks. In particular, the forth and fifth examples illustrate that it is possible for our algorithm to successfully handle some extremely difficult cases, where the blurry images are very unclear such that human eyes are unable to recognize their contents.
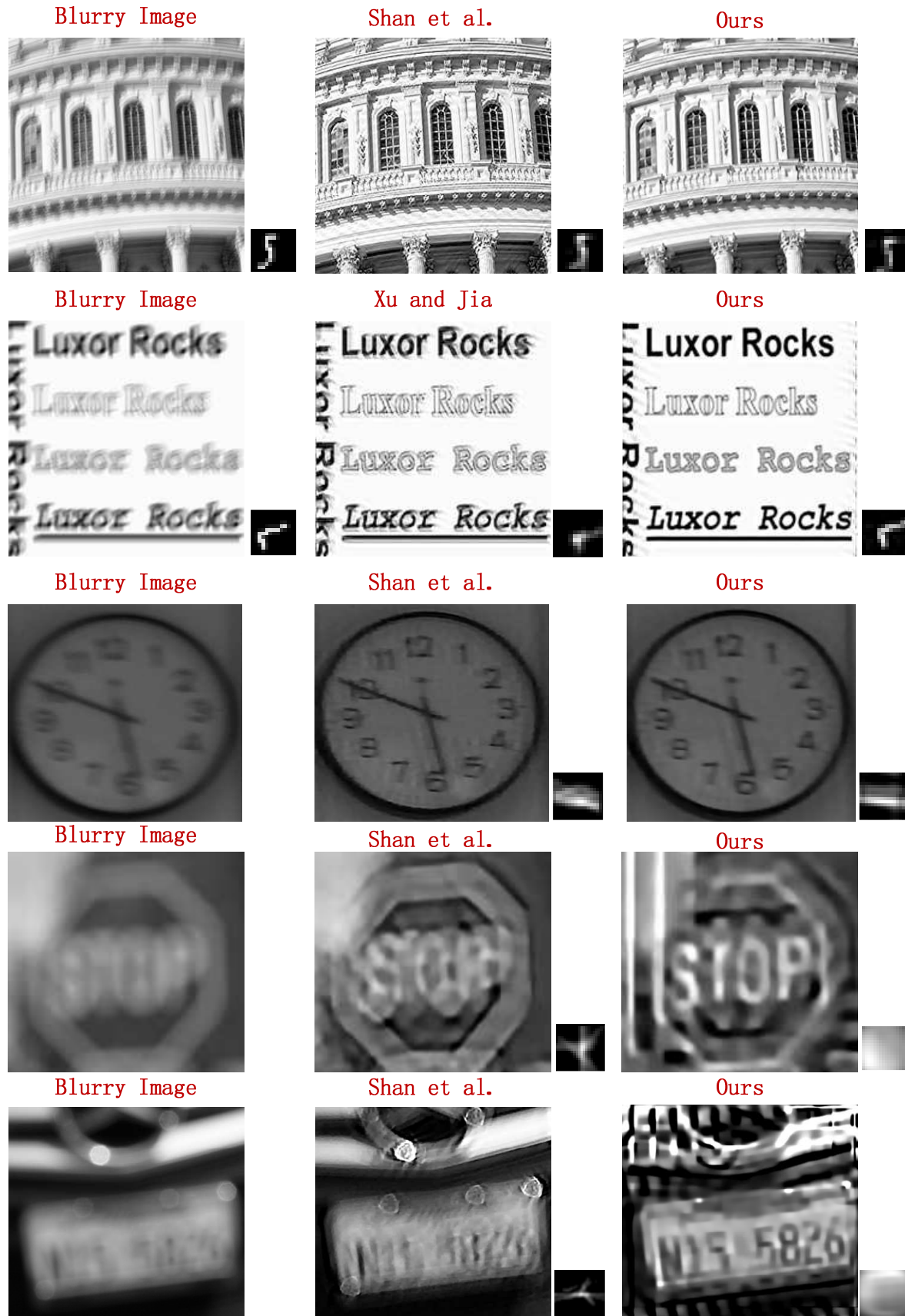
Fig. 4. **Comparison results, using two synthetical and three real examples.** Left: The observed blurry image. Middle: The best result among Fergus et al. [19], Shan et al. [18], Cho and Lee [17], Xu and Jia [13], and Cho et al. [10]. Right: Our result. For these five examples, the parameter α is set as 5050, 1050, 70, 4000, and 27000, respectively; the kernel sizes are, 13×13, 13×13, 17×17, 17×17, and 31×31, respectively. The parameters of the baselines are also manually tuned to best.
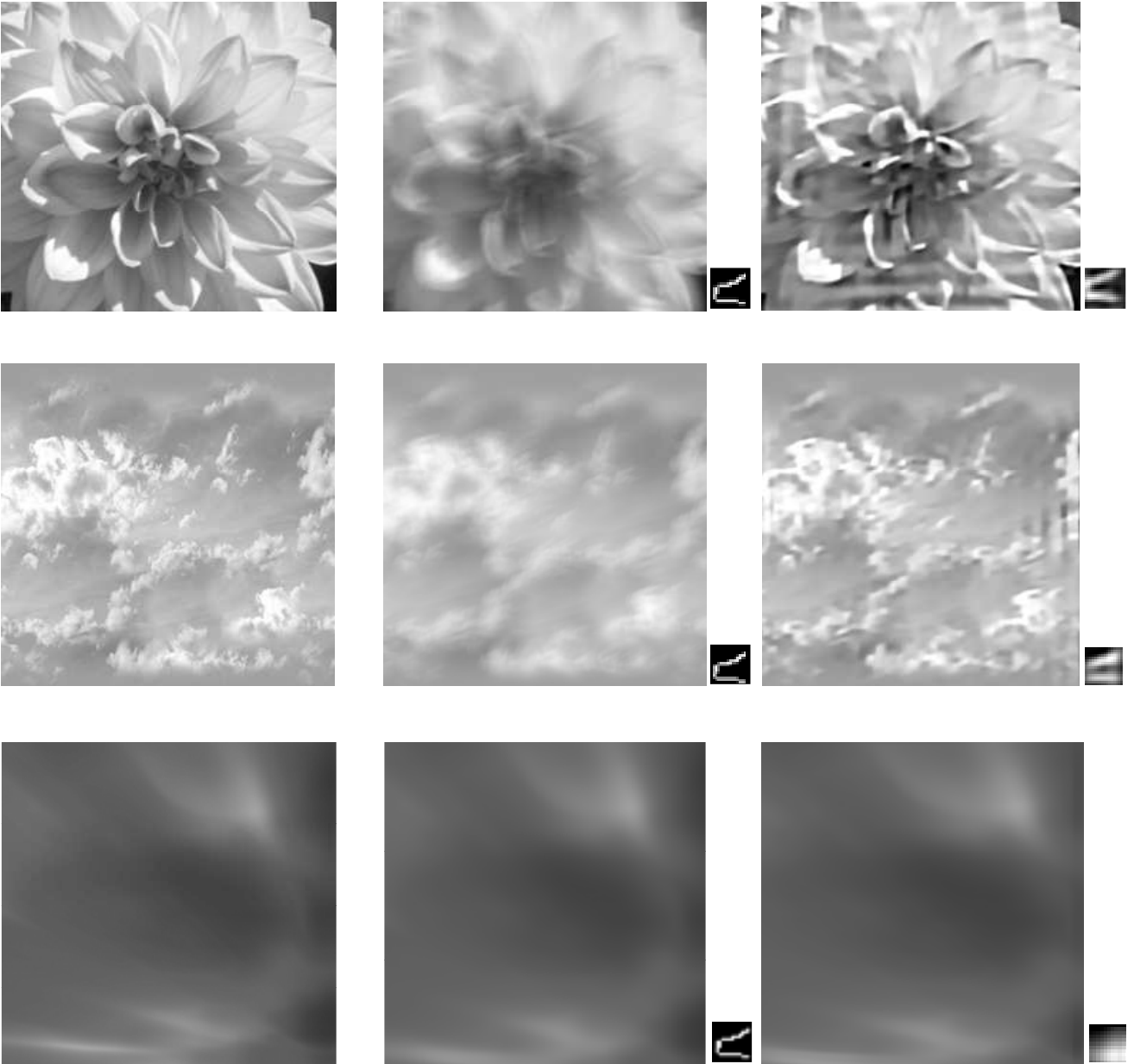
Fig. 5. **Some results obtained from the images with few edges.** Left: The original sharp image. Middle: The blurred version of the sharp image. Right: The deblurred version produced by our algorithm.

B. Diagnosis

A basic assumption of our approach is that the edges of a sharp image are different before and after blurring. So our algorithm is less effective while the images contain fewer edges, as shown in Figure 5. Particularly, the third example of Figure 5 shows that the "true" blur kernel cannot be recovered if the original image has very few edges. Nevetheless, it is not very critical to consider such images as there is almost no difference between their sharp and blurry versions.

Note that the forth example of Figure 4 may have non-uniform blur kernels, since the deblurred result favors the characters "O,P" over "S,T". This is true, as illustrated in Figure 6. While the blur is uniform within the whole image, the varying of α (within certain range) should not change the shape of the estimate of the blur kernel.

IV. Conclusions and Discussions

By studying how the spectrum of an image as a convolution operator changes before and after blurring, we have derived a convex regularizer on the blur kernel that depends on the given blurry image. For the blind deconvolution problem, we show that this convex regularizer is an effective prior that can deal with various types of realistic and challenging blurs (e.g., a combination of Gaussian and motion blurs). Both theoretical and experimental results have verified the validity

Fig. 6. **Non-uniform blur kernels.** Left: The observed blurry image. Middle: The deblurred version produced by our algorithm with $\alpha = 1000$. Right: The deblurred version produced by our algorithm with $\alpha = 4000$.

and effectiveness of the proposed prior. Notice that our regularizer harnesses certain necessary conditions on the blur kernel, but not sufficient. So it is entirely possible there might exist other, potentially more effective image-dependent regularizers for the blur kernel.

Theorem 2.1 and the empirical results in Figure 3 even suggest that in theory, for certain class of sharp images, it might be possible to solve the blind deconvolution problem by avoiding the iterative minimization altogether. Namely, the problem might be solvable by using only two computational stable procedures: Firstly estimate the blur kernel without estimating the sharp image at all; and then recover the sharp image by performing non-blind deconvolution. However, we leave such investigations for future.

APPENDIX

Proof: Let $\mathcal{T}_{u,v}(\cdot)$ denote the transformation operator that shifts a 2D function $I(\cdot,\cdot)$ from $I(x,y)$ to $I(x-u, y-v)$. By the convexity of $f(\cdot)$,

$$
\begin{aligned}
f(I \otimes K) &= f(\sum_{u,v} K(u,v)\mathcal{T}_{u,v}(I)) \\
&\leq \sum_{u,v} K(u,v)f(\mathcal{T}_{u,v}(I)) \\
&= f(I),
\end{aligned}
$$

where the last equality is due to $f(\mathcal{T}_{u,v}(I)) = f(I)$. This is naturally satisfied, as $\mathcal{T}_{u,v}(I)$ and I refer to the same image. ∎

Proof: Using the "min-max" half of the Courant-Fisher theorem, we have that

$$
\begin{aligned}
\sigma_i^{\mathcal{L}}(B) &= \min_{\substack{Y_j, \\ j=1,\cdots,i-1.}} \max_{\substack{\|X\|_F=1, \\ \langle X,Y_j\rangle=0, \\ j=1,\cdots,i-1.}} \|\mathcal{L}(B) \otimes X\|_F \\
&\leq \min_{\substack{Y_j=\kappa_j^{\mathcal{L}}(I), \\ j=1,\cdots,i-1.}} \max_{\substack{\|X\|_F=1, \\ \langle X,Y_j\rangle=0, \\ j=1,\cdots,i-1.}} \|\mathcal{L}(B) \otimes X\|_F \\
&= \max_{\substack{\|X\|_F=1, \\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0, \\ j=1,\cdots,i-1.}} \|\mathcal{L}(B) \otimes X\|_F.
\end{aligned}
$$

By (2), we have $\|\mathcal{L}(B) \otimes X\|_F \leq \|\mathcal{L}(I) \otimes X\|_F$ and thus

$$
\begin{aligned}
\sigma_i^{\mathcal{L}}(B) &\leq \max_{\substack{\|X\|_F=1, \\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0, \\ j=1,\cdots,i-1.}} \|\mathcal{L}(B) \otimes X\|_F \\
&\leq \max_{\substack{\|X\|_F=1, \\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0, \\ j=1,\cdots,i-1.}} \|\mathcal{L}(I) \otimes X\|_F \\
&= \sigma_i^{\mathcal{L}}(I),
\end{aligned}
$$

where the last equality is due to the definition of the convolution eigenvalues. ∎

The proof of Theorem 2.1 needs to use the following lemma.

Lemma A.1: If $B \neq 0$, then the Hessian matrix H defined by (13) is positive definite and

$$h^{\mathcal{L}(B)}(X) \geq \frac{s_1 s_2}{(\sigma_{max}^{\mathcal{L}}(B))^2} \|X\|_F^2, \forall X \in \mathbb{R}^{m_1 \times m_2}.$$

Proof: It could be calculated that

$$h^{\mathcal{L}(B)}(X) = \sum_{i=1}^{s_1 s_2} \frac{\|X \otimes \kappa_i^{\mathcal{L}}(B)\|_F^2}{(\sigma_i^{\mathcal{L}}(B))^2} \geq \frac{1}{(\sigma_{max}^{\mathcal{L}}(B))^2} \sum_{i=1}^{s_1 s_2} \|X \otimes \kappa_i^{\mathcal{L}}(B)\|_F^2 \doteq \frac{(\nu(X))^T H_1 \nu(X)}{(\sigma_{max}^{\mathcal{L}}(B))^2},$$

where $H_1 = \sum_{i=1}^{s_1 s_2} (\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B)))^T \mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))$. The Toeplitz matrix $\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))$ is a linear operator of $\kappa_i^{\mathcal{L}}(B)$ and can be explicitly written as

$$\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B)) \;=\; [(\nu(\kappa_i^{\mathcal{L}}(B)))^T \Omega_1; (\nu(\kappa_i^{\mathcal{L}}(B)))^T \Omega_2; \cdots],$$

where $\{\Omega_j\}_{j=1}^{(m_1+s_1-1)(m_2+s_2-1)}$ are a set of binary matrices and satisfy $\sum_j (\Omega_j)^T \Omega_j = s_1 s_2 \mathcal{I}$, and $\mathcal{I}$ is the $m_1 \times m_2$ identity matrix. Hence,

$$\begin{aligned}
H_1 &= \sum_{i=1}^{s_1 s_2} \sum_{j=1}^{(m_1+s_1-1)(m_2+s_2-1)} (\Omega_j)^T \nu(\kappa_i^{\mathcal{L}}(B))(\nu(\kappa_i^{\mathcal{L}}(B)))^T \Omega_j \\
&= \sum_{j=1}^{(m_1+s_1-1)(m_2+s_2-1)} (\Omega_j)^T (\sum_{i=1}^{s_1 s_2} \nu(\kappa_i^{\mathcal{L}}(B))(\nu(\kappa_i^{\mathcal{L}}(B)))^T) \Omega_j \\
&= \sum_{j=1}^{(m_1+s_1-1)(m_2+s_2-1)} (\Omega_j)^T \Omega_j, \\
&= s_1 s_2 \mathcal{I},
\end{aligned}$$

which simply leads to $h^{\mathcal{L}(B)}(X) \geq s_1 s_2 \|X\|_F^2 / (\sigma_{max}^{\mathcal{L}}(B))^2$. ∎

Proof: **of Theorem 2.1** By (11) and the convexity of $h^{\mathcal{L}(B)}(\cdot)$,

$$\begin{aligned}
h^{\mathcal{L}(B)}(K_0 - \hat{K}_0) &\leq h^{\mathcal{L}(B)}(K_0) + h^{\mathcal{L}(B)}(\hat{K}_0) \\
&\leq 2 h^{\mathcal{L}(B)}(K_0) \\
&\leq \frac{2 s_1 s_2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}.
\end{aligned}$$

By Lemma C.1, we also have

$$h^{\mathcal{L}(B)}(K_0 - \hat{K}_0) \geq \frac{s_1 s_2}{(\sigma_{max}^{\mathcal{L}}(B))^2} \|K_0 - \hat{K}_0\|_F^2.$$

Hence, $\|K_0 - \hat{K}_0\|_F^2 \leq 2(\sigma_{max}^{\mathcal{L}}(B))^2 / (\sigma_{min}^{\mathcal{L}}(I_0))^2$. ∎

Proof: Since $\hat{K}_0$ is deterministic, we have

$$\begin{aligned}
\|K_0' - K_0''\|_F &= \|(K_0' - \hat{K}_0) + (\hat{K}_0 - K_0'')\|_F \\
&\leq \|K_0' - \hat{K}_0\|_F + \|\hat{K}_0 - K_0''\|_F \\
&\leq \sqrt{2} \frac{\sigma_{max}^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0')} + \sqrt{2} \frac{\sigma_{max}^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0'')} \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

∎

We first prove the following lemma.

Lemma A.2: For any two matrices X and Y, we have

$$\|X \otimes Y\|_F \;\leq\; \|X\|_F \|Y\|_1,$$

where $\|\cdot\|$ is the ℓ_1-norm of a matrix.

Proof: Decompose Y into the sum of its negative part (denoted as Y^-) and nonnegative part (denoted as Y^+). By (2) and the convexity of matrix norms,

$$\begin{aligned}
\|X \otimes Y\|_F &= \|X \otimes Y^+ + X \otimes Y^-\|_F \\
&\leq \|X \otimes Y^+\|_F + \|X \otimes Y^-\|_F \\
&\leq \|X\|_F \|Y^+\|_1 + \|X\|_F \|Y^-\|_1 \\
&= \|X\|_F \|Y\|_1.
\end{aligned}$$

Proof: **Of Theorem 2.2** In noisy case, the inequality (10) need be changed to

$$
\begin{aligned}
\|K_0 \otimes \kappa_i^{\mathcal{L}}(B)\|_F &\leq \frac{\sigma_i^{\mathcal{L}}(B) + \|\mathcal{L}(N) \otimes \kappa_i^{\mathcal{L}}(B)\|_F}{\sigma_{min}^{\mathcal{L}}(I_0)} \\
&\leq \frac{\sigma_i^{\mathcal{L}}(B) + \epsilon\|\kappa_i^{\mathcal{L}}(B)\|_1}{\sigma_{min}^{\mathcal{L}}(I_0)} \\
&\leq \frac{\sigma_i^{\mathcal{L}}(B) + \sqrt{s_1 s_2}\epsilon}{\sigma_{min}^{\mathcal{L}}(I_0)}.
\end{aligned}
$$

Thus the inequality (11) becomes

$$
\begin{aligned}
h^{\mathcal{L}(B)}(K_0) &\leq \sum_i \frac{(1 + \frac{\sqrt{s_1 s_2}\epsilon}{\sigma_i^{\mathcal{L}}(B)})^2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2} \\
&\leq \frac{s_1 s_2 (1 + \frac{\sqrt{s_1 s_2}\epsilon}{\sigma_{min}^{\mathcal{L}}(B)})^2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}.
\end{aligned}
$$

Then by following the proof procedure of Theorem 2.1, this theorem can be proven.

REFERENCES

[1] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, pp. 55–59, 1972.
[2] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, 1988.
[3] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, 1996.
[4] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *NIPS*, 2009, pp. 1033–1041.
[5] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *CVPR*, 2009.
[6] N. Joshi, R. Szeliski, and D. J. Kriegman, "Psf estimation using sharp edge prediction," in *CVPR*, 2008.
[7] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, 2011.
[8] A. C. Likas and N. P. Galatsanos, "A variational approach for bayesian blind image deconvolution," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2222–2233, 2004.
[9] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
[10] T. S. Cho, S. Paris, B. K. P. Horn, and W. T. Freeman, "Blur kernel estimation using the radon transform," in *CVPR*, 2011, pp. 241–248.
[11] J. Jia, "Single image motion deblurring using transparency," in *CVPR*, 2007.
[12] A. Levin, "Blind motion deblurring using image statistics," in *NIPS*, 2006, pp. 841–848.
[13] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *ECCV*, 2010, pp. 157–170.
[14] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
[15] C. Jordan, "Sur la série de fourier," *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, vol. 92, pp. 228–230, 1881.
[16] Y.-L. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1723–1730, 2000.
[17] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, 2009.
[18] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, 2008.
[19] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.
[20] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 874–883, 2005.